# Making Randomness Public in Unbounded-round Information Complexity

Alexander Kozachinskiy

Lomonosov Moscow State University

**Abstract.** We prove a version of a "Reverse Newman Theorem" in information complexity: every private-coin communication protocol with information complexity $I$ and communication complexity $C$ can be converted into a public-coin protocol with the same behavior so that it's information complexity does not exceed $O\left(\sqrt{IC}\right)$. "Same behavior" means that the transcripts of these two protocols are identically distributed on each pair of inputs. Such a conversion was previously known only for one-way protocols. Our result provides a new proof for the best-known compression theorem in Information Complexity.

**Keywords:** We would like to encourage you to list your keywords within the abstract section

## 1 Introduction

Information complexity of a communication protocol $\pi$, denoted by $IC_\mu(\pi)$, is the amount of information Alice and Bob reveal to each other about their inputs while running $\pi$ under the assumption that their input pairs are distributed according $\mu$. Information complexity is used foremost in studying the Direct-Sum problem. Let us start with appropriate definitions.

Fix a small constant $\epsilon$. Suppose that we are given a function $f : \mathcal{X} \times \mathcal{Y} \to \{0,1\}$ and probability distribution $\mu$ on the set $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}$ is the set of Alice's inputs and $\mathcal{Y}$ is the set of Bob's inputs. The deterministic distributional complexity $D^\mu_\epsilon(f)$ is defined as

$$D^\mu_\epsilon(f) = \min_\pi CC(\pi),$$

where $CC(\pi)$ stands for the worst case communication complexity (i.e. the height) of $\pi$ and minimum is taken over all deterministic communication protocols $\pi$ which compute $f(x,y)$ on a random input pair, distribute according to $\mu$, with error probability at most $\epsilon$. Now imagine, that Alice and Bob have to compute $n$ copies of $f$ in parallel: Alice receives $n$ input $x$'s, $x_1, \ldots, x_n$ and Bob $n$ input $y$'s, $y_1, \ldots, y_n$, where the pairs $(x_i, y_i)$ are independent on each other and distributed according to $\mu$. In other words, they have to compute the function $f^n : (\mathcal{X} \times \mathcal{Y})^n \to \{0,1\}^n$ with input pairs distributed according to probability distribution $\mu^n$ on the set $(\mathcal{X} \times \mathcal{Y})^n$, which are defined as follows:

$$f^n\left((x_1, y_1), \ldots, (x_n, y_n)\right) = \left(f(x_1, y_1), \ldots, f(x_n, y_n)\right),$$

$$\mu^n((x_1,y_1),\ldots,(x_n,y_n)) = \mu(x_1,y_1)\times\ldots\times\mu(x_n,y_n).$$

This function has also its distributional communication complexity. However we are interested not in protocols computing $f^n$ with error probability at most $\epsilon$ with respect to $\mu^n$, but rather in protocols that compute each coordinate of $f^n$ with error probability at most $\epsilon$. That is, we consider deterministic communication protocols $\pi$ which output $n$ bits $\pi_1(x,y),\ldots,\pi_n(x,y)$ such that for every $i$ the following holds: $\mu^n\{(x,y)\,|\,\pi_i(x,y)\neq f(x_i,y_i)\}\leq\epsilon$. Then we consider the value

$$D_\epsilon^{n,\mu^n}(f^n) = \min_\pi CC(\pi),$$

where minimum ranges over all such protocols.

The definitions imply that $D_\epsilon^{n,\mu^n}(f^n)\leq nD_\epsilon^\mu(f)$ (apply the protocol witnessing $D_\epsilon^\mu(f)$ to compute each coordinate of $f^n$). The Direct-Sum question asks how close are $D_\epsilon^{n,\mu^n}(f^n)$ and $nD_\epsilon^\mu(f)$.

In an attempt to prove the opposite inequality $D_\epsilon^{n,\mu^n}(f^n)\geq nD_\epsilon^\mu(f)$ we can start with converting the protocol $\pi$ witnessing $D_\epsilon^{n,\mu^n}(f^n)$ into a randomized protocol $\tau$ using the technique described in [2]. $\tau$ computes $f$ with error probability at most $\epsilon$ (probability is taken with respect to the product distribution of $\mu$ and the distribution over the inner randomness of the protocol). Also $\tau$ satisfies: $IC_\mu(\tau)\leq CC(\pi)/n$, $CC(\tau)\leq CC(\pi)$. Assume now that any randomized protocol with communication complexity $C$, information complexity $I$ and error probability $\epsilon$ can be converted into a randomized protocol with communication complexity $\phi(I,C,\epsilon,\delta)$ computing the same function with error probability $\delta$. Here $\phi(I,C,\epsilon,\delta)$ is a certain function. Applying this conversion to the protocol $\tau$ we would obtain a randomized protocol with communication complexity

$$\phi\left(\frac{D_\epsilon^{n,\mu^n}(f^n)}{n}, D_\epsilon^{n,\mu^n}(f^n),\epsilon,\delta\right)$$

computing $f$ with error probability $\delta$. Using Yao's principle we then can convert that randomized protocol to a deterministic one with the same communication complexity and error probability.

Thus we are interested in "compression theorems" of the following form

| For every randomized protocol $\alpha$ which computes a function $g$ over the distribution $\mu$ with error probability $\epsilon$ there exists randomized protocol $\alpha'$ which computes $g$ over distribution $\mu$ with error probability $\delta$ such that $CC(\alpha')\leq\phi(IC_\mu(\alpha),CC(\alpha),\epsilon,\delta)$ |
|---|

**Fig. 1.** Compression statement for $\phi$

There are several compression theorems. The first one was proved in [1]:

**Theorem 1.** *Compression statement holds for* $\phi(I,C,\epsilon,\delta)=O\left(\sqrt{IC}\frac{\log(C/\rho)}{\rho}\right)$, *where* $\rho=\delta-\epsilon$.

This compression theorem implies that that

$$D^{\mu}_{\epsilon+\rho}(f) = O\left(\frac{D^{n,\mu^n}_{\epsilon}(f^n)\log(D^{n,\mu^n}_{\epsilon}(f^n)/\rho)}{\rho\sqrt{n}}\right).$$

In the above discussion we assumed that randomized protocols are allowed to use both private and public randomness. For protocols that use only public randomness there is a better compression theorem

**Theorem 2 ([4], [7]).** *Compression statement holds for public-coin protocols with* $\phi(I, C, \epsilon, \delta) = O(I\frac{\log(C/\rho)}{\rho})$*, where* $\rho = \delta - \epsilon$*.*

Unfortunately the randomized protocol $\tau$ mentioned above uses both public and private coins. Thus to benefit this theorem we have to convert the protocol $\tau$ into a protocol that uses public randomness only. It should be noted here that for Information Complexity private coins are more powerful than public coins. In contrast, for Communication Complexity the situation is the opposite: public coins are more powerful than private coins, but not very much: by Newman's theorem [6] every public coin randomized protocol can be converted to a private coin protocol at the expense of increasing the error probability by $\delta$ and communication complexity by $O(\log(n/\delta))$ (for any $\delta$ and for inputs of length $n$). We need a "reverse Newman theorem" for Information complexity, that is, a theorem stating that every private-coin protocol $\tau$ can be converted to a public-coin protocol $\tau'$ at the expense of increasing slightly the error probability and information complexity. Notice that we cannot covert $\tau$ to $\tau'$ just making private randomness publicly known. For example, assume that according to $\tau$ Alice sends to Bob the bit-wise XOR of her input $x$ and privately chosen random string $r$. Bob obtains no information about Alice's input from that message. However if $r$ is chosen publicly then Bob gets to know Alice's input.

We say that two protocols are *distributional-equivalent* if they are defined on the same input space $\mathcal{X} \times \mathcal{Y}$ and for every $(x, y) \in \mathcal{X} \times \mathcal{Y}$ their transcripts, conditioned on $(x, y)$, have the same probability distribution. Our contribution is the following

**Theorem 3.** *For every private-coin protocol* $\pi$ *there exists a public-coin protocol* $\tau$ *which is distributional-equivalent to* $\pi$ *(in particular,* $CC(\tau) = CC(\pi)$*) such that for every distribution* $\mu$ *the following holds:*

$$IC_{\mu}(\tau) = O\left(\sqrt{IC_{\mu}(\pi)CC(\pi)}\right).$$

*The constant hidden in O-notation is an absolute constant.*

Previously better conversions were known but only for bounded-round protocols. Namely, [4] establishes the conversion $IC_{\mu}(\tau) = IC_{\mu}(\pi) + O(\log(nl))$ for protocols running in constant number of rounds. Here $n$ is the length of input and $l$ is the length of randomness. And [3] proves a tight upper bound for one-way protocols: $IC_{\mu}(\tau) \leq IC_{\mu}(\pi) + \log IC_{\mu}(\pi) + O(1)$. In both results $\mu$ denotes

arbitrary probability distribution, $\pi$ denotes the given private-coin communication protocol $\pi$ and $\tau$ the constructed public-coin communication protocol $\tau$, which is distributional-equivalent to $\pi$ and does not depend on $\mu$.

Our result provides a new proof of Theorem 1: given a protocol $\alpha$ with communication complexity $C$ and information complexity $I$ we first convert it into a public-coin protocol with information complexity $O(\sqrt{IC})$. The communication complexity does not change, as the new protocol has the same distribution over transcripts than the original one. Then we apply Theorem 2 to the resulting public-coin protocol.

Notice that Theorem 1 (as well as any other compression theorem) implies a "reverse Newman theorem": every private-coin protocol $\tau$ with information complexity $I$, communication complexity $C$ and error probability $\epsilon$ can be converted to a public-coin protocol $\tau'$ with information complexity $O(\sqrt{IC}\log C)$ and error probability, for example, $2\epsilon$. Indeed, information complexity of any public-coin protocol does not exceed its communication complexity and we can consider the protocol existing by Theorem 1 as public-coin protocol (recall that for communication complexity public coins are at least as powerful as private coins). However the bound $O(\sqrt{IC}\log C)$ obtained in this way is $\log C$ larger than our bound. Besides the resulting public-coin protocol is not distributional-equivalent to the original one.

Our technique is not novice. The key fact is the relation between the statistical distance between Alice's and Bob's distributions of each bit sent in the protocol and the information revealed by sending that bit. This relation is established using Pinsker's inequality. It is worth to note that in the original proof of Theorem 1 the same idea is used to estimate the error probability of the converted protocol.

## 2 Preliminaries

Base 2 logarithms are denoted by log and natural logarithms by ln.

### 2.1 Information Theory

We use the standard notion of Shannon entropy; if $X$ is a random variable taking values in the set $\mathcal{X}$, then:

$$H(X) = \sum_{x\in\mathcal{X}} \Pr[X = x] \log\left(\frac{1}{\Pr[X = x]}\right).$$

By definition $0 \log 0 = 0$.

Assume that $X, Y$ are jointly distributed random variables. Then the conditional Shannon entropy $H(X|Y)$ is defined as $H(X|Y) = E_{y\leftarrow Y} H(X|Y = y)$. Here $X|Y = y$ denotes the random variable whose distribution is equal to the distribution of $X$ conditioned on the event $Y = y$ and $E_{y\leftarrow Y}$ stands for the

expectation over $y$ with respect to the marginal distribution of $Y$. It is easy to show that

$$H(X|Y) = H(X,Y) - H(Y).$$

Mutual information between jointly distributed random variables is defined as follows:

$$I(X:Y) = H(X) - H(X|Y).$$

Mutual information is symmetric: $I(X:Y) = I(Y:X)$; this follows from the above equality $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$.

For a triple $X, Y, Z$ of jointly distributed random variables we can consider conditional mutual information defined in a similar way:

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z).$$

Here $H(X|Y,Z)$ is an abbreviation for $H(X|(Y,Z))$. Entropy and the mutual information satisfy the chain rule:

**Proposition 1 (Chain Rule)**

$$H(X_1,\ldots,X_n) = H(X_1) + \sum_{i=2}^{n} H(X_i|X_1,\ldots,X_{i-1}),$$

$$I(X_1,\ldots,X_n:Y) = I(X_1:Y) + \sum_{i=2}^{n} I(X_i:Y|X_1,\ldots,X_{i-1}).$$

Chain rule holds also for conditional entropy and conditional mutual information.

Let $P$, $Q$ denote probability distributions on a finite set $W$. We consider two quantities that measure dissimilarity between $P$ and $Q$: *total variation*, or *statistical difference*:

$$\delta(P,Q) = \sup_{A\subset W} |P\{A\} - Q\{A\}|,$$

and the *information divergence*, or *Kullback-Leibler divergence*:

$$D_{KL}(P||Q) = \sum_{w\in W} P(w)\log\left(\frac{P(w)}{Q(w)}\right).$$

We will use the following well-known inequality:

**Proposition 2 (Pinkser's inequality)**

$$\delta(P,Q) \le \sqrt{\frac{D_{KL}(P||Q)}{2}}.$$

Mutual information between two joint distributed random variables can be expressed in terms of Kullback-Leibler divergence.

**Proposition 3** *If $Q$ is the distribution of $Y$ and $P_x$ is the distribution of $Y$ conditioned on the event $X = x$, then*

$$I(X : Y) = E_{x \leftarrow X} D_{KL}(P_x || Q).$$

When $\alpha$ is a real number between 0 and 1 we use denote by $H(\alpha)$ the entropy of a random variable $\xi$ with two possible values $\{w_1, w_2\}$ such that $\Pr[\xi = w_1] = \alpha$:

$$H(\alpha) = \alpha \log\left(\frac{1}{\alpha}\right) + (1-\alpha)\log\left(\frac{1}{1-\alpha}\right).$$

We will use the following fact:

**Fact 1** *If $\alpha \leq \frac{1}{2}$, then $H(\alpha) \leq 2\alpha \log\left(\frac{1}{\alpha}\right)$*

*Proof.* It is sufficient to show that $(1-\alpha)\log\left(\frac{1}{1-\alpha}\right) \leq \alpha \log\left(\frac{1}{\alpha}\right)$ for all $\alpha \leq \frac{1}{2}$. To this end consider the function $f(\alpha) = \alpha \log\left(\frac{1}{\alpha}\right) - (1-\alpha)\log\left(\frac{1}{1-\alpha}\right)$. The derivative of this function equals

$$f'(\alpha) = \frac{1}{\ln(2)}\left(\ln\left(\frac{1}{\alpha(1-\alpha)}\right) - 2\right).$$

The equation $\alpha(1-\alpha) = 1/e^2$ has two different roots $\alpha_0 < \alpha_1$ and the derivative is negative for all $\alpha$ between the roots and positive outside. For $\alpha = 1/2$ the derivative is negative thus the function $f$ increases on $[0, \alpha_0]$, and decreases on $[\alpha_0, \frac{1}{2}]$. Since $f(0) = f(\frac{1}{2}) = 0$ this implies that $f(\alpha) \geq 0$ for all $\alpha \in [0, \frac{1}{2}]$. □

### 2.2 Communication Protocols

*The definition of a deterministic communication protocol.* A deterministic protocol to compute a function $f : \mathcal{X} \times \mathcal{Y} \to \mathcal{Z}$ is specified by a functions $\delta : \{0,1\}^* \to \{A, B\} \cup \mathcal{Z}$, indicating the turn to communicate and the output when communication is over; and functions $p : \mathcal{X} \times \mathcal{A} \to \{0,1\}$, $q : \mathcal{Y} \times \mathcal{B} \to \{0,1\}$, which instruct Alice and Bob how to communicate. Here $\mathcal{A} = \{s \in \{0,1\}^* \mid \delta(s) = A\}$ and $\mathcal{B} = \{s \in \{0,1\}^* \mid \delta(s) = B\}$.

The figure 2 how the protocol specified by $\delta, p, q$ is performed.

1. Alice receives $x \in \mathcal{X}$, Bob receives $y \in \mathcal{Y}$; they add some bits to the string $s$, called the *transcript*. At the start of the protocol the transcript is empty: $s = \lambda$;
2. If $s \in \mathcal{A}$, Alice sends the bit $b = p(x, s)$ and then Alice and Bob append $b$ to $s$;
3. If $s \in \mathcal{B}$, Bob acts in a similar way, using the function $q$ instead of $p$;
4. If $s \in \mathcal{O} = \{s \in \{0,1\}^* \mid \delta(s) \in \mathcal{Z}\}$, then Alice and Bob output $\delta(s)$ and terminate.

**Fig. 2.** Running a deterministic communication protocol.

The length of the transcript at the end of the protocol is called the *communication length* of the protocol for that pair. The maximal communication length

(over all input pairs) is called the *communication complexity* of the protocol $\pi$, denoted by $CC(\pi)$.

We say that a deterministic protocol *computes* the function $f$ if for all input pairs $x, y$ the protocol outputs $f(x, y)$.

*The definition of a randomized communication protocol.* A randomized protocol is defined similarly to deterministic protocols. However this time functions $p$ and $q$ are "random functions". That is, Alice has a random variable $R^A$ and Bob has a random variable $R^B$ with outcomes in some sets $U, V$; the function $p$ maps $\mathcal{X} \times \{0,1\}^* \times U$ to $\{0,1\}$ and the function $q$ maps $\mathcal{Y} \times \{0,1\}^* \times V$ to $\{0,1\}$. At the start protocol Alice and Bob sample $R^A$ and $R^B$ and then both act deterministicaly, using the functions $p(\cdot, \cdot, R^A)$ and $q(\cdot, \cdot, R^B)$, respectively.

The transcript occurred in a randomized protocol depends not only on the inputs of Alice and Bob but also on their randomness: for each input pair $x, y$ the transcript is a random variable. The maximum length of the transcript that can occur with positive probability for a specific input pair is called *communication length* of the protocol for that pair (it may be infinite). The maximal communication length (over all input pairs) is called the *communication complexity* of the protocol $\pi$, denoted by $CC(\pi)$.

We say that a randomized protocol $\pi$ *computes* the function $f$ with error probability $\epsilon$ if for all input pairs $x, y$ with probability at least $1 - \epsilon$ it happens that $\pi$ outputs $f(x, y)$ on input pair $(x, y)$.

Whether a randomized protocol is private-coin or public coin depends on the joint probability distribution of the random variables $R^A, R^B$. If the random variables $R^A$ and $R^B$ are independent then the protocol is called *private-coin*. In a private-coin protocol each party gets know the bits sent by the other party but does not know the randomness that has caused sending those bits.

If $R^A = R^B$, that is, Alice and Bob use the same randomness, then the protocol is called *public-coin*. The common value of $R^A$ and $R^B$ is denoted by $R$ and is called *shared* or *public randomness*. One can consider also an intermediate case: $R^A$ and $R^B$ are dependent but do not coincide. We will not need such protocols in this paper.

Every private-coin protocol $\pi$ with randomness $R^A, R^B$ can be converted into a public-coin protocol with shared randomness equal to the pair $(R^A, R^B)$. The communication complexity and error probability of this public-coin protocol are the same as those of the original private-coin protocol. Moreover, the resulting protocol is distributionally equivalent to the original one. A similar conversion in the other direction is impossible. This means that with respect to communication complexity public coins are more powerful than private coins.

### 2.3 Information Complexity

The *information complexity of a randomized protocol* $\pi$ with respect to a probability distribution $\mu$ over input pairs is defined by the formula

$$\begin{aligned} IC_\mu(\pi) &= I(X : \Pi, R^B \mid Y) + I(Y : \Pi, R^A \mid X) \\ &= I(X : \Pi \mid R^B, Y) + I(Y : \Pi \mid R^A, X). \end{aligned}$$

Here $X, Y, \Pi, R^A, R^B$ denote jointly distributed random variables, where $R^A, R^B$ are Alice's and Bob's randomness, $(X, Y)$ is a random pair of inputs drawn according to $\mu$ and $\Pi$ is the transcript of the protocol for those $X, Y, R^A, R^B$. So $\Pi$ is a deterministic function of the other variables. The pair of variables $(R^A, R^B)$ is independent from the pair $(X, Y)$.

In this formula, $I(X : \Pi, R^B \mid Y)$ accounts of the information about Alice's input revealed to Bob by running the protocol and $I(Y : \Pi, R^A \mid X)$ accounts of the information about Bob's input revealed to Alice by running the protocol The two expressions for information complexity are the same, and moreover,

$$I(X : \Pi, R^B \mid Y) = I(X : \Pi \mid Y, R^B), \qquad I(Y : \Pi, R^A \mid X) = I(Y : \Pi \mid X, R^A).$$

Indeed, $X$ and $R^B$ are independent conditional to $Y$ and $Y$ and $R^A$ are independent conditional to $X$.

**Lemma 1.** *For private-coin protocols, we have* $I(X : \Pi | R^B, Y) = I(X : \Pi | Y)$ *and* $I(Y : \Pi | R^A, X) = I(Y : \Pi | X)$

*Proof.* Indeed, the difference between the former two quantities can be written as

$$I(X : \Pi | R^B, Y) - I(X : \Pi | Y) = I(X : R^B | \Pi, Y) - I(X : R^B | Y). \qquad (1)$$

This equality can be verified by expressing all its terms through unconditional entropy. Both terms in the right hand side of (1) are zeros. Indeed, by definition $X$ and $R^B$ are independent conditional to $Y$.

Also $X$ and $R^B$ are independent conditional to $Y, \Pi$. This is not obvious and follows from the rectangle property of deterministic protocols: for each $s \in \mathcal{O}$ the set of all pairs of inputs that produce the transcript $s$ is a combinatorial rectangle, that is, a Cartesian product of some sets (see [5]). This implies that for any randomized protocol the set of all pairs $\langle (x, r^A), (y, r^B) \rangle$ that produce a certain transcript $s$ is a combinatorial rectangle, too.

Fix $s$ and $y$. By definition, the random variables $(X, R^A)$ and $(Y, R^B)$ are independent conditional to the event $Y = y$. The condition "$X, R^A, Y, R^B$ produce $s$" means that $\langle (X, R^A), (Y, R^B) \rangle$ belongs to some rectangle $P \times Q$. Adding such a condition to the condition $Y = y$ does not make $(X, R^A)$ and $(Y, R^B)$ dependent. Therefore, $(X, R^A)$ and $(Y, R^B)$ and hence $X$ and $R^B$ are independent conditional to $(\Pi, Y)$. □

For public-coin protocols the formula of informational complexity becomes

$$IC_\mu(\pi) = I(X : \Pi, R | Y) + I(Y : \Pi, R | X),$$

where $R$ stands for the shared randomness.

**Lemma 2.** *For public-coin protocols,*

$$IC_\mu(\pi) = H(\Pi | R, Y) + H(\Pi | R, X).$$

*Proof.* Indeed, we have

$$\begin{aligned} I(X : \Pi, R|Y) &= H(\Pi, R|Y) - H(\Pi, R|X, Y) \\ &= H(\Pi|R, Y) + H(R|Y) - H(\Pi|R, X, Y) - H(R|X, Y). \end{aligned}$$

We have $H(\Pi|R, X, Y) = 0$, since $\Pi$ is determined by $R, X, Y$. Furthermore, $H(R|Y) = H(R|X, Y) = H(R)$ since $R$ is independent from the pair $(X, Y)$. Hence $I(X : \Pi, R|Y) = H(\Pi|R, Y)$. In a similar way we can prove that $I(Y : \Pi, R|X) = H(\Pi|R, X)$. □

If we apply the conversion from private-coin to public-coin protocols described in the end of the previous section, the resulting protocol may have much larger information complexity then the original protocol. For example, it happens for the protocol where Alice sends to Bob the bit-wise XOR of her input and her private random string. The purpose of the present paper is to construct a more smart conversion, such that the resulting public-coin protocol has the least known information complexity (for many-round protocols).

## 3 Simulation of One-Bit Protocols

Let us start with proving Theorem 3 for one-way protocols of depth 1. That is, for protocols with only one bit sent, say by Alice.

We are given a private-coin protocol $\pi$, where a single bit is sent and it is sent by Alice. Such protocol is specified by a function $p : \mathcal{X} \times U \to \{0, 1\}$, a random variable $R^A$ with values in $U$ and a function $\delta : \{0, 1\} \to \mathcal{Z}$. For input $x$ and private randomness $r \in U$ Alice sends the bit $p(x, r)$. Let $P_x$ denote the distribution of the random variable $p(x, \cdot)$ that is $P_x(i) = \Pr[p(x, R^A) = i]$ for $i = 0, 1$.

We define public-coin protocol $\tau$ as follows:

1. Alice receives $x \in \mathcal{X}$;
2. Alice and Bob publicly sample $R$ uniformly in $[0, 1]$;
3. Alice sends $B(x, R)$, where $B(x, R) = 0$ if $R < P_x(0)$ and $B(x, R) = 1$ otherwise.

It is clear that for every $x$ Alice's message $B$ is distributed according to $P_x$. Hence $\tau$ is distributional-equivalent to $\pi$.

Assume now that we are given a probability distribution $\mu$ on the set $\mathcal{X} \times \mathcal{Y}$ which defines random variable $(X, Y)$. We have to show that for some constant $D$ it holds

$$IC_\mu(\tau) \leq D\sqrt{IC_\mu(\pi)} \tag{2}$$

Notice that in both protocols $\pi, \tau$ no information about Bob's input is revealed to Alice. By Lemmas 1 and 2 we have $IC_\mu(\pi) = I(X : B|Y)$ and $IC_\mu(\tau) = H(B|R, Y)$. Assume first that Bob's input is fixed. That is, there is a $y_0$ with such that $Y = y_0$ with probability 1. Then in the formulas for information complexity we can drop the condition $Y$.

We have to relate Information Complexity of $\tau$ to that of $\pi$. The former equals

$$IC_\mu(\tau) = H(B|R) = \int_0^1 H(B|R=t)dt,$$

where the random variable $B$ denotes the bit sent by Alice, i.e., $B = B(X,R)$, and $B|R=t$ denotes the distribution of $B(X,t)$.

The latter equals $IC_\mu(\pi) = I(X:B) = E_{x\leftarrow\mu} D_{KL}(P_x||Q)$, where $Q$ denotes the distribution of $B$ (see Proposition 3).

Thus we have to show that

$$\int_0^1 H(B|R=t)dt = O(\sqrt{E_{x\leftarrow\mu} D_{KL}(P_x||Q)}).$$

By Pinsker's inequality (Proposition 2) we have: $(\delta(P_x,Q))^2 \leq D_{KL}(P_x||Q)/2$ and hence it suffices to prove that

$$\int_0^1 H(B|R=t)dt = O(\sqrt{V}), \tag{3}$$

where

$$V = E_{x\leftarrow\mu}\delta^2(P_x,Q).$$

Consider the set

$$\Omega = \left\{ t \in [0,1] \mid |t - Q(0)| > \sqrt{2V} \right\}.$$

It is clear that $\Pr[R \notin \Omega] \leq 2\sqrt{2V}$ hence

$$\int_{[0,1]\setminus\Omega} H(B|R=t)dt \leq 2\sqrt{2V}. \tag{4}$$

Fix $t \in \Omega$. We claim that either $\mu\{x \mid B(x,t)=0\}$ or $\mu\{x \mid B(x,t)=1\}$ is at most $V/(t-Q(0))^2$. Assume first that $t < Q(0) - \sqrt{2V}$. Then $B(x,t)=1$ implies $P_x(0) \leq t$ hence

$$\delta(P_x,Q) = |P_x(0) - Q(0)| \geq |t - Q(0)|.$$

By Markov's inequality

$$\mu\{x \,|\, B(x,t)=1\} \leq \frac{V}{(t-Q(0))^2}.$$

The case $t > Q(0) + \sqrt{2V}$ is entirely similar, in this case $\mu\{x \mid B(x,t) = 0\} \leq V/(t-Q(0))^2$. By fact 1, and because $V/(t-Q(0))^2 \leq 1/2$, we get:

$$\begin{aligned}
\int_{\Omega} H(B|R=t)dt &\leq \int_{\Omega} H\left(\frac{V}{(t-Q(0))^2}\right) dt \\
&\leq 2\int_{\Omega} \frac{V}{(t-Q(0))^2} \log\left(\frac{(t-Q(0))^2}{V}\right) dt \\
&\leq 2\sqrt{V}\int_{\Omega} \frac{V}{(t-Q(0))^2} \log\left(\frac{(t-Q(0))^2}{V}\right) d\frac{(t-Q(0))}{\sqrt{V}} \\
&\leq 2\sqrt{V}\int_{|y|>\sqrt{2}} \frac{\log y^2}{y^2} dy = O(\sqrt{V}).
\end{aligned}$$

The last equality holds, as the integral $\int_{|y|>\sqrt{2}} \frac{\log y^2}{y^2} dy$ converges. Thus we have proved (3) and (2).

It remains to prove the inequality (2) in the general case (when Bob's input is not fixed). In this case we may observe that both left hand side and right hand side of (2) are linear combinations of conditional entropies with $Y$ in condition. If we fix any $y \in Y$ and replace $Y$ in the condition by $Y = y$, then the inequality (2) becomes valid, as we just have proved. Averaging over $y$ proves (2) as it is.

We conclude this section by presenting a randomized protocol from [3] showing that our bound for one-bit protocols is tight. Assume that Alice receives 0 or 1 with equal probabilities and then sends one bit to Bob, which is equal to her input bit with probability $\frac{1}{2}+\epsilon$ and differs from it with probability $\frac{1}{2}-\epsilon$. One can show that for every public-coin implementation of this protocol, with probability $2\epsilon$ Bob learns Alice's input hence the information complexity of the protocol is at least $2\epsilon$. At the same time a simple calculation shows that if random bits are private, then information complexity drops to $\Theta(\epsilon^2)$.

## 4 The Generalization to All Protocols

In this section we extend the result of the previous section to all protocols.

*Proof (of theorem 3).* Assume that $\pi$ is an arbitrary private-coin communication protocol, defined by the functions $\delta, p, q$ and random variables $R^A, R^B$. First we convert $\pi$ to another private-coin protocol $\pi'$ in which each bit is sent using a fresh randomness (independent on randomness used to send previous bits). The protocol $\pi'$ will be distributional-equivalent to $\pi$ and hence will have the same information complexity.

The private-coin protocol $\pi'$ works as follows:

1. Alice receives $x \in \mathcal{X}$, Bob receives $y \in \mathcal{Y}$; they let $s = \lambda$. Until $s \in \mathcal{O}$ they perform the following items 2 and 3.

2. If $s \in \mathcal{A}$, $|s| = k$, then Alice reads a real $r_k \in [0,1]$ from the its private random source and sends $p'(x, s, r_k)$ which equals 0 if

$$r_k < \Pr[p(x, s, R^A) = 0 \mid E_s]$$

and 1 otherwise. Here $E_s$ denotes the intersection over $i \leq |s|$ with $s_{1\ldots i-1} \in \mathcal{A}$ of the events

$$p(x, s_{1\ldots i-1}, R^A) = s_i.$$

The set $E_s$ depends only on $s$ and $x$, thus Alice is able to find $\Pr[p(x, s, R^A) = 0 \mid E_s]$. The sent bit is then appended to $s$.

3. If $s \in \mathcal{B}$, Bob acts in a similar way;
4. If $s \in \mathcal{O}$, Alice and Bob output $\delta(s)$ and terminate.

By construction $\pi'$ is distributional-equivalent to $\pi$. Assume that we are given a probability distribution $\mu$ on $\mathcal{X} \times \mathcal{Y}$ which defines random variables $X, Y$. By Lemma 1, for private-coin protocols the information complexity depends only on the distribution of the triple $X, Y, \Pi$, the information complexities of $\pi$ and $\pi'$ coincide.

The public-coin protocol $\tau$ is obtained from $\pi'$ be just assuming that all the strings $r_k$ are read from the shared random source. Notice that $\tau$ does not depend on $\mu$. By construction $\tau$ is distributional-equivalent to $\pi$.

We have to relate information complexity of the private-coin protocol $\pi'$ to that of the constructed public-coin protocol $\tau$.

Set $N = CC(\pi)$ and let $\Pi = \Pi_1 \ldots \Pi_N$ denote the transcript of $\pi'$. W.l.o.g. we may assume that for all inputs and all randomness the number of sent bits equals $N$ (Alice can send fixed bits when output is decided). Set $\Pi_{<k} = \Pi_1 \ldots \Pi_{k-1}$.

By chain rule (Proposition 1), applied to protocol $\pi'$ we have:

$$\begin{aligned} IC_\mu(\pi') &= I(X : \Pi|Y) + I(Y : \Pi|X) \\ &= \sum_{k=1}^{N} I(X : \Pi_k|Y, \Pi_{<k}) + I(Y : \Pi_k|X, \Pi_{<k}) \\ &= \sum_{k=1}^{N} I_k, \end{aligned}$$

where $I_k = I(X : \Pi_k|Y, \Pi_{<k}) + I(Y : \Pi_k|X, \Pi_{<k})$.

We claim that

$$IC_\mu(\tau) \leq D\sqrt{I_1} + \cdots + D\sqrt{I_N}.$$

To prove the claim note that by Lemma 2 we have $IC_\mu(\tau) = H(\Pi|R,Y) + H(\Pi|R,X)$ where $R = (r_0, \ldots, r_{N-1})$. By chain rule we get:

$$\begin{aligned} IC_\mu(\tau) &= H(\Pi|R,Y) + H(\Pi|R,X) \\ &= \sum_{k=1}^{N} H(\Pi_k|R,Y,\Pi_{<k}) + H(\Pi_k|R,X,\Pi_{<k}) \\ &= \sum_{k=1}^{N} I'_k, \end{aligned}$$

where

$$I'_k = H(\Pi_k|R,Y,\Pi_{<k}) + H(\Pi_k|R,X,\Pi_{<k}).$$

Thus to prove the claim it suffices to show that $I'_k \le D\sqrt{I_k}$.

Indeed, $I_k$ is the average over all $s \in \{0,1\}^{k-1}$ of $I(X : \Pi_k|Y,\Pi_{<k} = s) + I(Y : \Pi_k|X,\Pi_{<k} = s)$. For every fixed $s$ consider the one-round private-coin protocol $\pi'_s$, in which Alice (if $s \in \mathcal{A}$, with obvious changes when $s \in \mathcal{B}$) samples a real $r_{k-1} \in [0,1]$ and sends $p'(x,s,r_{k-1})$ to Bob. The quantity $I(X : \Pi_k|Y,\Pi_{<k} = s) + I(Y : \Pi_k|X,\Pi_{<k} = s)$ is then the information complexity of $\pi'_s$ with respect to the distribution $X,Y|\Pi_{<k} = s$.

The conversion of the previous section applied to the protocol $\pi'_s$ yields the public-coin protocol that is the same as $\pi'_s$ except that now $r_{k-1}$ is read from the random source. From the previous section it follows that

$$\begin{aligned} H(\Pi_k|r_k,Y,\Pi_{<k} = s) + H(\Pi_k|r_k,X,\Pi_{<k} = s) \\ \le D\sqrt{I(X : \Pi_k|Y,\Pi_{<k} = s) + I(Y : \Pi_k|X,\Pi_{<k} = s)}, \end{aligned}$$

and hence

$$\begin{aligned} H(\Pi_k|R,Y,\Pi_{<k} = s) + H(\Pi_k|R,X,\Pi_{<k} = s) \\ \le D\sqrt{I(X : \Pi_k|Y,\Pi_{<k} = s) + I(Y : \Pi_k|X,\Pi_{<k} = s)}. \end{aligned}$$

The value $I'_k$ is the expectation over $s$ of the left hand side of the last inequality. Similarly the value $I_k$ is the expectation of the expression under the radical in the right hand side. As the square root function is concave this implies

$$I'_k \le D\sqrt{I_k}$$

Using Cauchy–Schwarz inequality we conclude

$$\begin{aligned} IC_\mu(\tau) &= I'_1 + \ldots + I'_N \\ &\le D\left(\sqrt{I_1} + \ldots + \sqrt{I_N}\right) \\ &\le D\sqrt{(I_1 + \ldots + I_N)N} = D\sqrt{IC_\mu(\pi)CC(\pi)}.\Box \end{aligned}$$

## References


1. Barak, B., Braverman, M., Chen, X., and Rao, A. How to compress interactive communication. *SIAM Journal on Computing 42*, 3 (2013), 1327–1363.
2. Braverman, M. Interactive information complexity. In *Proceedings of the 44th symposium on Theory of Computing* (2012), ACM, pp. 505–524.
3. Braverman, M., and Garg, A. Public vs private coin in bounded-round information. In *Automata, Languages, and Programming*. Springer, 2014, pp. 502–513.
4. Brody, J., Buhrman, H., Koucky, M., Loff, B., Speelman, F., and Vereshchagin, N. Towards a reverse newman's theorem in interactive information complexity. In *Computational Complexity (CCC), 2013 IEEE Conference on* (2013), IEEE, pp. 24–33.
5. Kushilevitz, E., and Nisan, N. *Communication complexity*. Cambridge university press, 2006.
6. Newman, I. Private vs. common random bits in communication complexity. *Information processing letters 39*, 2 (1991), 67–71.
7. Pankratov, D. *Direct sum questions in classical communication complexity*. PhD thesis, Masters thesis, University of Chicago, 2012.